



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/662,602	09/15/2003	Chang-Ning Huang	M61.12-0514	2369

27366 7590 12/14/2009  
WESTMAN CHAMPLIN (MICROSOFT CORPORATION)  
SUITE 1400  
900 SECOND AVENUE SOUTH  
MINNEAPOLIS, MN 55402

EXAMINER
----------

SAINT CYR, LEONARD

ART UNIT	PAPER NUMBER
----------	--------------

2626

MAIL DATE	DELIVERY MODE
-----------	---------------

12/14/2009

PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.



## DETAILED ACTION

### *Response to Arguments*

1. Applicant's arguments filed 08/03/09 have been fully considered but they are not persuasive.

Applicants argue that the statistics generated in the Palmer reference have nothing to do with quantifying a level of precision with which word type indications have been applied (Amendment, page 8).

The examiner disagrees, since Palmer disclose **"In scoring segmentation using algorithm, recall is defined as the percentage of actual words from the hand-segmented text identified in the corresponding positions in the text, while precision is defined as the percentage of identified words which are also in the same positions in the hand-segmented text. The NMSU segmenter consists of an initial approximation followed by a sequence of iterative refinements...to recognize idiomatic expressions, derived words, Chinese person names, and foreign proper names.** It will be interesting to determine the contribution of each of these to the segmentation accuracy as well as the retrieval score. **Similarly, it may be helpful to use frequency-based phrase building, that is, segmentation based on character n-gram occurrences in the collection"** (page 176, col.2; page 177; col.2).

Applicants argue that neither Palmer et al., nor Dien et al., morphological output as claimed that includes an affixation type indication of a plural affix applied to a noun (Amendment, page 9).

The examiner disagrees, since Dien et al., disclose “**those are the words morphologically derived...**Similar to English, **there appear also prefixes and suffixes, which are however much simpler in the morphology of Vietnamese. Therefore, we apply further morphological analysis to easily identify this class of words.** The critical point here is to determine the weight of these derived words” [section 4.1.5].

### ***Claim Rejections - 35 USC § 102***

2. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

3. Claim 47 is rejected under 35 U.S.C. 102(b) as being anticipated by Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997).

As per claims 47, and 57, Palmer et al., teach a computer-implemented method for evaluating a word segmentation language model, comprising:

building the word segmentation language model based on an annotated corpus (“segmentation based on character bigrams”; page 175, col.2, paragraph 1; page 177, col.2, paragraph 5);

utilizing a computer processor that is a functional component of the computer to apply the language model to a test corpus of unsegmented text different from the

Art Unit: 2626

annotated corpus so as to provide an output indicative of words in the test corpus and a word type indication for each word, the word type indication being any one of a plurality of word type indications (“**refinement steps attempt to recognize idiomatic expressions, derived words, Chinese person names, and foreign proper names**” those are considered as word type indications; pages 177, col.2, paragraph 3);

utilizing the processor to compare the word type indication for each word in the output of the language model with predefined word type indications of words of the test corpus; and utilizing the processor to automatically generate a quantitative value that represents a level of precision with which word type indications were applied in the output indicative of words in the test corpus (“precision is defined as the percentage of identified words which are also in the same positions in the hand-segmented text... **refinement steps attempt to recognize idiomatic expressions, derived words, Chinese person names, and foreign proper names**, determine the contribution of each of these steps to the segmentation accuracy”; page 176, col.2, paragraph 5; pages 177, col.2, paragraph 3);

wherein generating comprises generating based on how frequently (“**it may be helpful to use frequency-based phrase building, that is, segmentation based on character n-gram occurrences**”) location name type indications (“proper names”) for words in the output match identical corresponding predefined location name indications assigned to the same words in the test corpus, and wherein generating comprises generating not based simply on the words in the output themselves but also generating based on a comparison involving the location name types assigned to words in the

Art Unit: 2626

output, each location name being a data descriptor that is separate and distinct from a word in the output to which it is assigned (**“precision is defined as the percentage of identified words which are also in the same positions in the hand-segmented text. The NMSU segmenter consists of an initial approximation followed by a sequence of iterative refinements...to recognize idiomatic expressions, derived words, Chinese person names, and foreign proper names”**; page 176, col.2; page 177; col.2).

***Claim Rejections - 35 USC § 103***

4. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.
5. Claims 55, 56 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Haizhou et al., (Chinese Word Segmentation, 1998).

As per claims 55, 56, Palmer et al., do not specifically teach that overlapping ambiguous string word type indications were applied in the output; organization name word type indications were applied in the output; covering ambiguous string word type indications were applied in the output.

Haizhou et al., teach As indicated by Liang[2], there are two cases of unexpected segmentation. One is overlapping ambiguity where a character could go either way to form two words, such as in example 1. Another is composition ambiguity where the subsegmentation is possible:...One can find that , and all are possible word entries,

Art Unit: 2626

thus both results are valid based on lexicon entries (page 215, section 3.2, paragraph 3).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made use overlapping ambiguous string word and covering ambiguous string word in precision scores as taught by Haizhou et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

6. Claims 63, 64, 67- 69, and 71 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Dien et al., (Vietnamese Word Segmentation, 2001).

As per claim 63, Palmer et al., teach a computer-implemented method for performing word segmentation, the method comprising:

receiving an input of unsegmented text; utilizing a computer processor that is a functional component of the computer to apply a language model so as to determine a segmentation of the unsegmented text (“segmentation based on character bigrams”; page 175, col.2, paragraph 1; page 177, col.2, paragraph 5);

identifying a morphologically derived word within the unsegmented text; and providing an output that includes the segmentation of the unsegmented text (“derived words...to the segmentation accuracy”; page 177, col.2, paragraph 3).

However, Palmer et al., do not specifically teach an indication of a combination of parts that form the morphologically derived word, the output also including an indication

Art Unit: 2626

of a part of speech for the combination of parts, and the output also including an indication that the morphological derived word demonstrates characteristics consistent with a morphological pattern of an affixation type.

Dien et al., teach that **those are the words morphologically derived**...Similar to English, **there appear also prefixes and suffixes, which are however much simpler in the morphology of Vietnamese. Therefore, we apply further morphological analysis to easily identify this class of words.** The critical point here is to determine the weight of these derived words [section 4.1.5].

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made use indication of a combination of parts that form derived words as taught by Dien et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

As per claim 68, Palmer et al., teach a computer-implemented method for performing word segmentation, the method comprising:

receiving an input of unsegmented text; utilizing a computer processor that is a functional component of the computer to apply a language model so as to determine a segmentation of the unsegmented text ("segmentation based on character bigrams"; page 175, col.2, paragraph 1; page 177, col.2, paragraph 5);

identifying a morphologically derived word within the unsegmented text; and providing an output that includes the segmentation of the unsegmented text ("derived words...to the segmentation accuracy"; page 177, col.2, paragraph 3).



However, Palmer et al., do not specifically teach an indication of a combination of parts that form the morphologically derived word, the output also including an indication of a part of speech for the combination of parts, and the output also including an indication that the morphological derived word demonstrates characteristics consistent with a morphological pattern of a reduplication type.

Dien et al., teach that no dictionary can be comprehensive enough with all these reduplicatives due to no exhaustive statistics. Here we make use of the rule of morpheme transformation in reduplicatives to identify them...**Those are the words morphologically derived...Therefore, we apply further morphological analysis to easily identify this class of words.** The critical point here is to determine the weight of these derived words [sections 4.1.4, and 4.1.5].

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made use indication of a combination of parts that form derived words as taught by Dien et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

As per claims 64, and 69, Palmer et al., further disclose the output also includes an indication of a named entity detected within the unsegmented text ("Chinese person names"; page 177, col.2, paragraph 3).

As per claim 67, Palmer et al., in view of Dien et al., further disclose that said indication that the morphologically derived word demonstrates characteristics consistent

Art Unit: 2626

with the morphological pattern of the affixation type is more specifically an indication that the morphological derived word demonstrates characteristics consistent with affixation of a plural affix to a noun (**“those are the words morphologically derived...Similar to English, there appear also prefixes and suffixes, which are however much simpler in the morphology of Vietnamese. Therefore, we apply further morphological analysis to easily identify this class of words”**; section 4.1.5).

As per claim 71, Palmer et al., in view of Dien et al., further disclose that said indication that the morphologically derived word demonstrates characteristics consistent with the morphological pattern of the reduplication type is more specifically an indication that the morphological derived word demonstrates characteristics consistent with transformation of an original word consisting of a pattern of characters into another word also consisting of the pattern of characters (“Here we make use of the rule of morpheme transformation in reduplicatives to identify them...**Those are the words morphologically derived...Therefore, we apply further morphological analysis to easily identify this class of words”**; sections 4.1.4, and 4.1.5).

7. Claims 65, and 70 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Dien et al., (Vietnamese Word Segmentation, 2001), and further in view of Guo et al., (US PAP 2002/0052901).

Art Unit: 2626

As per claims 65, and 70, Palmer et al., in view of Dien et al., do not specifically teach the output also includes an indication of a factoid detected within the unsegmented text.

Guo et al., teach that for Chinese language, the morphologic process includes the steps of: (1) segmenting sentences into words according to the system dictionary and the user-defined dictionaries; (2) identifying proper names (currently including person names, place names and person titles), domain terms, **numbers, measure words, and date expressions** (paragraph 29).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to identify factoids as taught by Guo et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

### ***Conclusion***

8. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any

Art Unit: 2626

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to LEONARD SAINT CYR whose telephone number is (571) 272-4247. The examiner can normally be reached on Mon- Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is (571)-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

LS  
12/03/09

Application/Control Number: 10/662,602

Page 12

Art Unit: 2626

/Richmond Dorvil/

Supervisory Patent Examiner, Art Unit 2626